

W42 - A Scalable Spatial Database System for Holding Digital Elevation Models

Martin Habermeyer, Ursula Marschalk, Achim Roth
German Aerospace Center (DLR)
German Remote Sensing Data Center (DFD)
82234 Wessling, Germany
(martin.habermeyer,ursula.marschalk,achim.roth)@dlr.de

Abstract—The design of a scalable system for holding spatial data in general and Digital Elevation Models (DEMs) in specific has to account for the characteristics of data from various application fields. The data can be heterogeneous in coverage, as well as in resolution, information content and quality. A database aiming at the representation of world-wide DEMs has to consider these differences in the design of the system with respect to the structure and the algorithms. The database system W42, which is presented in the work at hand, is a scalable spatial database system capable of holding, extracting, mosaicking, and fusing spatial data represented in raster- as well as in vector-format. Design aspects for this task can be specified as holding spatial data in unique data structures and providing unique access functions to the data. These are subject of this work as well as first experiences gained from the implementation of part of the extensions made for the TanDEM-X mission.

I. INTRODUCTION

The design of a scalable system for holding spatial data has to be flexible to account for the characteristics of the products from different application fields. In the context of DEMs the products from different sources can be heterogeneous in coverage, as well as in resolution, information content and quality. A database aiming at the representation of world-wide DEMs has to consider these differences in the design with respect to the structure and the algorithms.

The W42-DEM-Database is a scalable stand-alone database system capable of holding spatial data in a unique system architecture. It provides functions allowing import and export of spatial data in raster and vector-like structures. The purpose of this document is the clarification of the database design and the description of the provided functionality. First results from the implementation of the extensions made for the upcoming TanDEM-X mission are presented.

This work extends the papers and reports previously published about the database (e.g. [1], [2], [3]). The database served as and still is part of the geocoding system for radar missions: SRTM/X-SAR [4], TerraSAR-X [5]. It is also used in an operational processing chain for optical airborne data [6].

The focus in this paper is on design aspects with regard to the TanDEM-X mission, for which the database has been adopted to be a part of the Mosaicking and Calibration

Processor ([7]) and to support accompanying studies (e.g. [8]). Whereas in [3] the preliminary design of the database was discussed, the present work extends the design and adds first experiences from the implementation of the tracks and the import of the first datasets.

The document is structured as follows: Section II covers the System Design in a Four-Layer-Concept, the Directory Structure (section V), and describes the kind of data (Tiles, Chips, Vectors) that can be stored in the database system at the moment (see section III).

The data types are abstracted to data formats (section IV). Based on the data formats substantial routines are introduced for the manipulation of the data. They are explained in section VI, without a consideration of an organisational structure, only working on a file level without spatial semantics. The functions presented cover file format conversion, file manipulation, information about the data content, visualization of files, fusing the information provided by different layers and different formats, mosaicking them together, resampling routines to change their resolution.

For each data type (data tile, image chip, track) then the needed routines for import, export, update, inventory and maintenance to, from and for the database are specified and their impact on database components is discussed (section VII). These utilities for the manipulation of the database content are based on programs providing functionalities to handle raster- and vector data.

Two examples covering the visualization of the database content and the capability for information fusion round the picture of the functionality currently provided by the spatial database.

II. SYSTEM DESIGN

The database's structure is organized in four layers. The system is based on the "Physical Layer" represented by media mounted on workstations or PCs. The second layer, which is named "File System Layer" spans the directory tree over the "Physical Layer". It allows unique access to the data and is organized in projects. Inside of each project

the geographic coordinate system allows the separation into smaller categories. The access to the "File System Layer" is carried out by the third layer, the "Database Access Layer". This layer provides unique functions to import, export, update, backup and remove data from and to the database. It further allows the display of the information content of the database. These functions are realized by the operating system and by the fourth layer, the "File Access Layer".

The system has to be scalable to allow an easy expansion of all four layers. This can be assured as we show for each layer. In case of the "Physical Layer" the system can be expanded by the integration of further media to adjust the available memory capacity to more projects and/or finer resolutions. The "File System Layer" can be adopted to hold further projects and higher resolutions by the introduction of new directories serving as containers to hold the expanded data sources. The "Database Access Layer" can account for the increase of resolution and the availability by the provision of more sophisticated functions. The "File Access Layer" is completely open for an extension of the data available and the enhancement of the available resolutions. The scalability of the system enables the integration of the database into operational processors: at the moment it is serving as part of the geocoding-processor of the TerraSAR-X mission and as part of the generic processing chain for airborne hyperspectral data. It further will be utilized in satellite missions as an element of the Payload Ground Segment in the future. Currently the database is adopted for the use in the ground segment of the TanDEM-X mission.

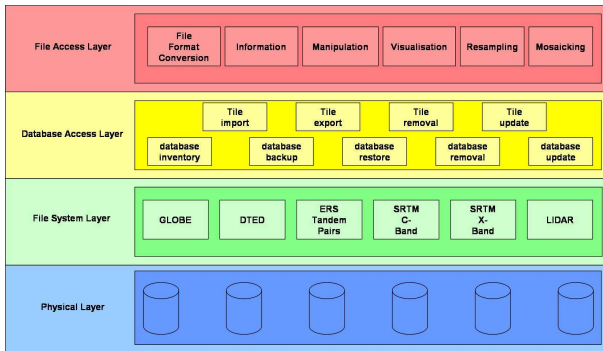


Figure 1. System Design

As the "File Access Layer" provides the functionality for the manipulation of the files in the database it is discussed here. There are seven classes in the "File Access Layer" providing the following basic functions:

- file format conversion routines for the support of different in- and output formats to keep the system open for future duties, which can go beyond a database system for DEMs,
- modules to display the information content of the data to have a quick tool allowing rudimentary quality control,

- tools to allow the manipulation of file content for the adoption of the data to other tasks,
- visualization utilities to support a quick overview of data capabilities,
- methods to resample the data to different resolutions or coordinate systems to open the system's data to a wider range of applications,
- utilities to allow the mosaicking of data: these comprise the combination of data from different sources as well as from different areas,
- utilities to support the fusion of data, where more than one height model for an area is available. Resolution and quality of the candidates have to be taken into account and carefully combined to arrange a product that combines the information content of different layers.

These utilities of the "File Access Layer" form a base of modules that can be combined to new modules.

The functionality of the "Database Access Layer" is provided by the combination of functions from the "File Access Layer" and the operating system. These elements can be composed to new modules (see fig. 2).

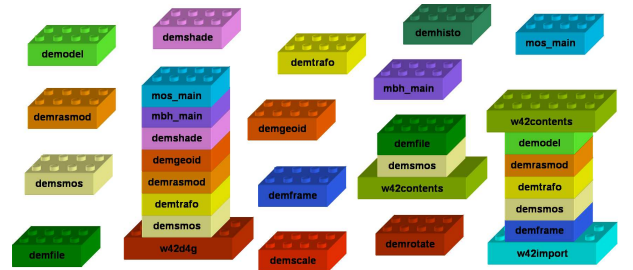


Figure 2. Modular Design of Routines

III. DATA TYPES

Currently the support for two data types is implemented: Data Tiles and Tracks. The third datatype, Image Chips, is listed here for completeness, but remains to be implemented.

A. Data Tiles

Data Tiles contain spatial data representing areal phenomena. Their borders are congruent with the grid spanned by the geographic coordinate. Their resolution of the pixels is one of 30, 10, 3, 1, 0.3, 0.1, 0.03, 0.01 arc-seconds. They are stored in Sun Raster Format with DLR extension (see section IV) and organized as specified in section V-A.

B. Image Chips

Image Chips contain spatial data representing areal phenomena. Their borders can be incongruent with the grid spanned by the geographic coordinate system (see fig. 4). The resolution of the chips can range from 30 to 0.01 arc-seconds. Image chips are stored in Sun Raster Format with DLR extension (see section IV) and organized as specified in section V-B.

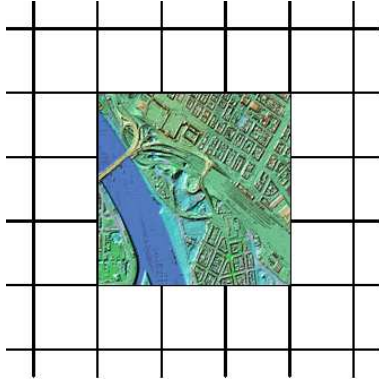


Figure 3. Data Tiles

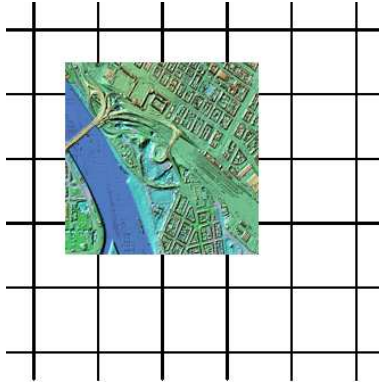


Figure 4. Image Chips

C. Tracks

Tracks contain spatial data representing line phenomena (see fig. 4).

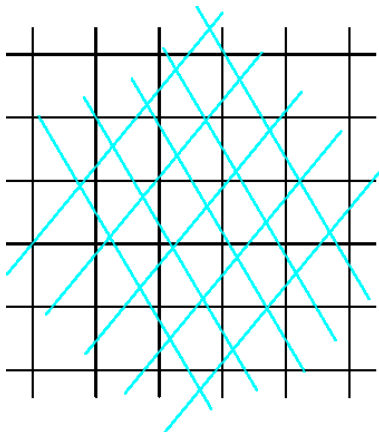


Figure 5. Tracks

IV. DATA FORMAT SPECIFICATION

For the representation of Image Chips and Data Tiles the Sun Raster File with DLR extension (SRF) [9] is foreseen.

The header of the files is held in Big-Endian byteorder format strictly to avoid confusion. For the purpose of enabling fast processing the files are kept in machine-specific byteorder after the export and prior to the import to the database. The DLR footer specifies statistics and the annotation, allowing a fast display of statistical information and quick addressing for Data Tiles and Image Chips.

The vector and point data is represented in ESRI's shapefile format [10].

V. ORGANISATION OF THE FILE SYSTEM

The database structure is based on a UNIX file system. The different sources of data stored is accounted for by separate branches (called projects in the following) in the directory tree.

Below the project level the grid spanned by the geographic coordinate system (fig. 6) is projected onto the directory tree of the file system in the order North (two digits ranging from 00 – 89) to South (two digits ranging from 01 – 90) and East (three digits ranging from 000 – 179) to West (three digits ranging from 001 – 180) in 1° steps. For resolutions finer than 1 arcsec the directory tree is further subdivided in 0.01 arcsec steps in latitudes and longitudes as well (in both directions represented by two digits ranging from 00 – 99). At each level a logfile is kept to keep track of changes made to the subdirectory (fig. 6).

The different data formats (Data Tiles, Image chips, and Tracks) sort into this file system in specific manners.

A. Data Tiles

DEMs are provided in $1^\circ \times 1^\circ$ degree tiles in geographic coordinates in Sun Raster Format. The various DEM sources plus additional data are organized as directory-trees from North to South and East to West in 1° steps. For resolutions below 1 arcsec the directory tree is further split up into 0.01 degree tiles. For performance reasons each tile is held in lower resolutions as well. This enables an incorporation of data with fine resolution (e.g. height models from laser scanners) as well as quick access to data only available in relatively coarse resolution (e.g. GLOBE data).

Data Tiles are provided in exact $1^\circ \times 1^\circ$ tiles geographic coordinates in Sun Raster Format with DLR extension. The data is zipped prior to storage. As mentioned before, the various DEM sources plus additional data are organized in directory-trees from North to South and East to West in 1° steps, and for resolutions below 1° in 0.01° steps (fig. 6). The finest resolution supported is 0.01 arcsec. For performance reasons each tile of a fine resolution is held in coarser resolutions as well. For an overview of the resolution of the stored data tiles check table I.

Each tile contains an overlap of 30 arcseconds to the South and the West to the adjacent tiles at the $1^\circ \times 1^\circ$ directory level and an 0.3 arcsecond overlap at the $0.01^\circ \times 0.01^\circ$ directory level.

For the upper left and the lower right coordinate the pixel

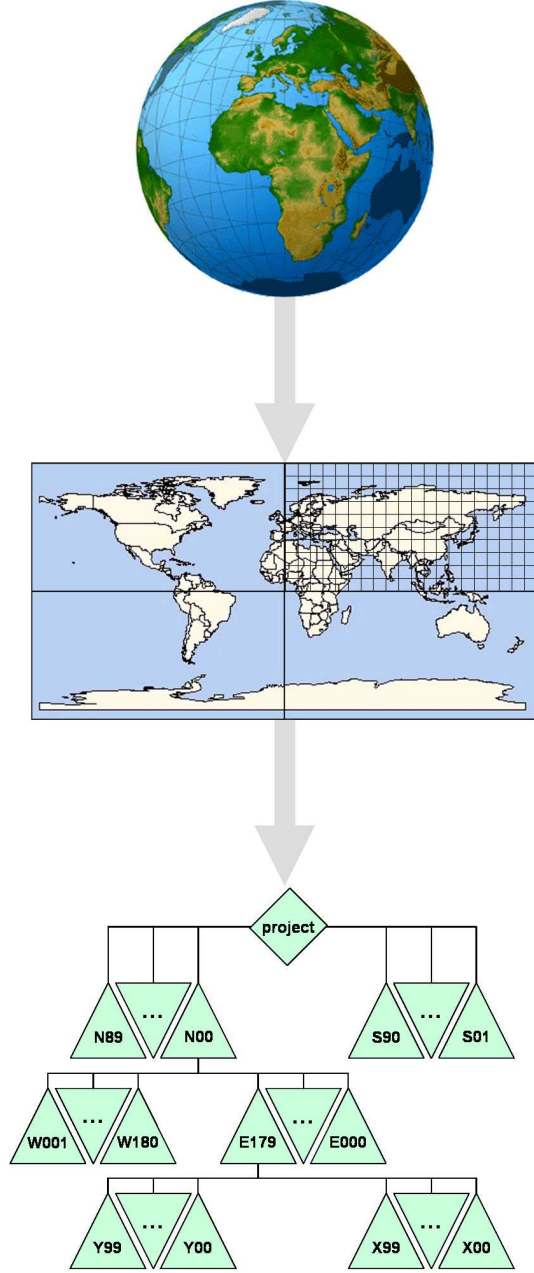


Figure 6. Projection of geographic coordinate system into database

centers are annotated. As lower right coordinate the first coordinate outside the image is specified. For changing resolutions the tile border is kept, which means that the coordinates change in the different layers as well.

At the moment the following different elevation data with a global coverage are stored as Data Tiles in the database:

- the GLOBE dataset [11] with a global coverage and a resolution of 1 km,
- data from the Shuttle-Radar-Topography-Mission (SRTM) [12]. SRTM mapped 58 million km^2 with DLR's X-band system and 119 km^2 of the land surface

Resolution (per pixel)	File Size (in pixels)	Directory Level
30 arcsec	121×121	$1^\circ \times 1^\circ$
10 arcsec	363×363	$1^\circ \times 1^\circ$
3 arcsec	1210×1210	$1^\circ \times 1^\circ$
1 arcsec	3630×3630	$1^\circ \times 1^\circ$
0.3 arcsec	121×121	$0.01^\circ \times 0.01^\circ$
0.1 arcsec	363×363	$0.01^\circ \times 0.01^\circ$
0.03 arcsec	1210×1210	$0.01^\circ \times 0.01^\circ$
0.01 arcsec	3630×3630	$0.01^\circ \times 0.01^\circ$

Table I
RESOLUTIONS OF DATA TILES PROVIDED IN THE DATABASE

Source / Format	Accuracy relative	Accuracy absolute	resolution (in arcsecs)
DTED1	20m	30m	1
DTED2	20m	30m	1
ERS	20m	30m	1
SRTM/C-Band	8m	16m	3
SRTM/C-Band	8m	16m	1
SRTM/X-Band	6m	16m	1
TanDEM-X	2m	10m	0.4

Table II
RESOLUTIONS AND ACCURACIES OF DIFFERENT SOURCES HELD IN THE DATABASE

with NASA's C-band system within the 56° Southern and 60° Northern latitudes,

- data resulting from repeat-pass-interferometric DEM production with the ERS-1 and ERS-2 satellites.
- a reference dataset used in the ground segment of the TerraSAR-X mission, composed of SRTM and GLOBE data.

As the nature of the data has an impact on the requirements for the algorithms and consequently the design of an elevation model database the characteristics of the DEMs provided by the different sensors/missions is summarized in table II. The accuracies to be provided by TanDEM-X are also listed.

B. Image Chips

Image chips are stored in the given directory structure (fig. 6). They are referenced by their center coordinate. Different from the data tiles they are only stored in their original resolution and not in the resolutions specified by table I. Their resolution in fact can be any resolution, and needn't fulfill any specification except for being specified in degrees or seconds. The directory structure has to be expanded by two components:

- an identifier file (only holding the actual identifier number) ensuring a unique addressing of the image chips,
- and a file containing a list where all image chips are tracked that are completely or in parts situated in the 1×1 degree tile to guarantee a quick search for tiles.

C. Tracks

Tracks are also stored in the directory structure shown in figure 6 and discussed in section V. The relevant information,

	Raster Data	Raster and Vector Data	Vector Data
Data Conversion	x		x
Data Manipulation	x		x
Data Information	x		x
Data Visualisation	x		x
Information Fusion	x	x	x
Mosaicking	x		x
Resampling	x	x	x

Table III
FILE ACCESS FUNCTIONS

including mandatory elevation data as well as additional optional information is kept in ESRI's [10] shapefile format. The data can be browsed for elevation data as well as for the additional information.

Currently the following sources are incorporated in the tracks-branch of the DEM-database:

- ICESat (Ice, Cloud and land Elevation Satellite) [13] laser altimetry data, providing multi-year elevation data from January 2003 up to now,
- selected tracks measured with a GPS.

Both data sources are used for the calibration and validation of the elevation data [8] provided by the TanDEM-X mission being launched in autumn 2009.

VI. FILE ACCESS ROUTINES

The functionalities provided for the handling of raster and vector data are displayed in table III. This comprises seven classes of basic functions:

- file format conversion routines for the support of different in- and output formats (currently support is given to SRF, GeoTiff, DTED format),
- routines to allow the manipulation of the file content,
- modules to display the information content of the data,
- visualization utilities, e.g. for the generation of color shaded presentations,
- algorithms to fuse overlaying tiles in different resolutions and qualities,
- utilities to allow the mosaicking of data, and
- methods to resample the data to different resolutions or coordinate systems,

VII. DATABASE ACCESS ROUTINES

For each datatype (tiles, chips, and tracks) unique routines have to be foreseen allowing the import, export, inventory and maintenance of the data. These are based on the File Access Routines described above (section VI) and the functionality provided by the operating system.

A. Import

For all data types exclusive write access is guaranteed by lock files. Log files are kept for a later reproduction of changes to the database. The inventory is created or updated for a quick

overview of the database's content.

Data Tiles have to be resampled to lower resolutions to provide a fast access to the needed accuracy later. In case of the Image Chips the chip list is updated, no resampling of the data is done. Tracks are partitioned to fit in the tiles where they are stored and sorted by their coordinates to allow fast access functions.

For each data type the components are zipped to save memory space.

B. Export

1) *Data Tiles*: The export routine for data tiles is composed of the following steps:

- get DEM tiles from the addressed project(s),
- mosaick DEM tiles to one file per project,
- cut out specified part for each project,
- fuse information from different layers to one file,
- apply coordinate transformation to target projection/coordinate system.

Figure 7 shows a color-shaded exported tile, where all of the above steps have been applied.

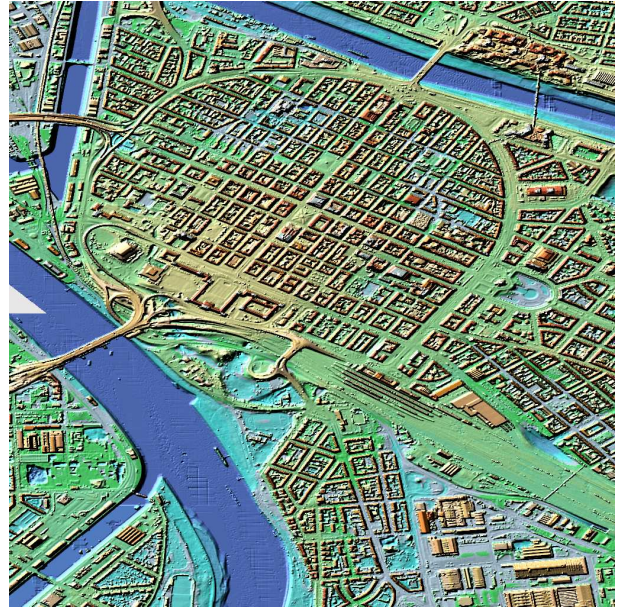


Figure 7. Exported Data Tiles, previously mosaicked, and resampled (LASER-DEM of the city of Mannheim, Germany)

2) *Image Chips*: The export routine for image chips accomplishes the following function: "for a given region of the earth's surface (specified by the corner coordinates) provide me with all image chips intersecting this region".

3) *Tracks*: The export routine for tracks accomplishes the following two functions:

- "for a given region of the earth's surface (specified by the corner coordinates) provide me with all points stored in tracks within this region".
- "for a given track provide me with all points lying in this region".

Both functions are realized by the use of the coordinates as well as the additional parameters. Currently the export to kml and shapefiles is supported (see fig. 8).

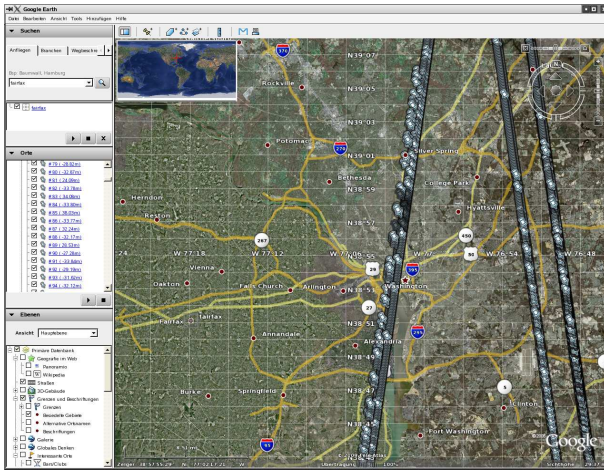


Figure 8. Exported Tracks as kml in a Google-Earth window

C. Update

Update routines for all data types have the following steps in common: creation and deleting of lock files providing exclusive file access, update of log files and inventory. Data tiles can be updated at any level. The coarser resolutions in this case have to be updated as well. The update of an image chip has to be interpreted as a replacement of an existing chip by a new chip. The identification of the chip to be updated is done by its unique identifier and has to be given by the operator. In this case the tile specified with its unique identifier is replaced. No update of the chip list and no incrementation of the currentIdentifier file is intended. For updating Tracks insertion and deletion of points into or from a Track have to be foreseen.

D. Inventory

For each file type an inventory function is foreseen that is capable of providing an overview of the data stored within (parts of) the project.

E. Maintenance

The maintenance module for each data type should be capable of providing the following functionality: undoing the last creation/update step, removing files from archives and inventory lists, erasing complete directory trees, backing up (parts of) the directory tree, copying (parts of) the directory tree. All of the above functions can also only be simulated, without an execution of the commands. In this case only listings are prepared instead.

VIII. CONCLUSIONS

The design of a scalable database system has been discussed. The concept for efficiently holding and accessing the data has been presented. It has been discussed how the different layers of the database can be adopted for further tasks. The data types supported at the moment have been specified.

An overview of the datasets currently held in the database, used as a DEM database was given. Two examples have shown that the data previously imported to the database can be exported to generic image processing formats.

For the preparation and support of the TanDEM-X Mosaicking and Calibration Processor another datatype remains to be implemented.

REFERENCES

- [1] *TerraSAR-X TMSP W42 DEM Database Interface Document*, DLR, Oberpfaffenhofen, 2006.
- [2] A. Roth, W. Knöpfle, G. Strunz, M. Lehner, and P. Reinartz, "Towards a global elevation product: combination of multi-source digital elevation models," in *Symposium on Geospatial Theory, Processing and Applications*, Ottawa, 2002.
- [3] M. Habermeyer, U. Marschalk, and A. Roth, "Digital elevation model database w42 - a scalable system for spatial data," in *The XXI Congress of The International Society for Photogrammetry and Remote Sensing*, vol. XXXVII, no. B1, 07 2008, pp. 1253 – 1258.
- [4] A. Roth, W. Knöpfle, B. Rabus, S. Gebhardt, and D. Scales, "GeMoS – a system for the geocoding and mosaicking of interferometric digital elevation models," in *Proceedings of IGARSS '99*, Hamburg, Germany, 1999.
- [5] M. Huber, B. Wessel, and A. Roth, "The TerraSAR-X orthorectification service and its benefit for land use applications," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2006.
- [6] M. Habermeyer, A. Müller, S. Holzwarth, R. Richter, R. Müller, K.-H. Seitz, P. Seifert, and P. Strobl, "Developing a fully automatic processing chain for the upcoming hyperspectral scanner ARES," in *Proceedings of the 3rd EARSeL Workshop on Imaging Spectroscopy*, M. Habermeyer, A. Müller, and S. Holzwarth, Eds., Herrsching, Germany, May 2003, pp. 82–87.
- [7] B. Wessel, U. Marschalk, A. Gruber, M. Huber, T. Hahmann, A. Roth, and M. Habermeyer, "Design of the DEM mosaicking and calibration processor for TanDEM-X," in *Proceedings of the 7th European Conference on Synthetic Aperture Radar*, 02.-05. June 2008, Friedrichshafen, Germany, 2008.
- [8] M. Huber, B. Wessel, D. Kosmann, A. Felbier, V. Schwieger, M. Habermeyer, A. Wendler, and A. Roth, "Ensuring globally the TanDEM-X height accuracy: analysis of the reference data sets ICESat, SRTM and KGPS-tracks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2009.
- [9] *File Formats used by DFD-Geocoding Software EGEO and GEMOS*, DLR, Oberpfaffenhofen, 2002.
- [10] *ESRI Shapefile Technical Description*, ESRI, Redlands, California, USA, 1998.
- [11] D. Hastings and P. Dunbar, "Development and assessment of the global one-km base elevation digital elevation model (GLOBE)," *International Archives of Photogrammetry and Remote Sensing*, vol. 32, no. 4, pp. 218–221, 1998.
- [12] C. Pessagno, "SRTM status report," in *Validation of DEMs from satellite data. Reports and proceedings of 7th Meeting of Mapping Sub Group*, I. Dowman, Ed. Gaithersburg, USA: CEOS WG Calibration and Validation, 2000.
- [13] H. Zwally, B. Schutz, W. Abdalati, J. Abshire, C. Bentley, A. Brenner, J. Bufton, J. Dezio, D. Hancock, D. Harding, T. Herring, B. Minster, K. Quinn, S. Palm, J. Spinhrine, and R. Thomas, "ICESat's laser measurements of polar ice, atmosphere, ocean, and land," *Journal of Geodynamics*, vol. 34, no. 3-4, pp. 405–445, 2002.